# Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/NO2005/000465

International filing date: 19 December 2005 (19.12.2005)

Document type: Certified copy of priority document

Document details: Country/Office: NO
Number: 20052535
Filing date: 27 May 2005 (27.05.2005)

Date of receipt at the International Bureau: 23 January 2006 (23.01.2006)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

3

KONGERIKET NORGE
The Kingdom of Norway

# Bekreftelse på patentsøknad nr
*Certification of patent application no*

**20052535**

Det bekreftes herved at vedheftede dokument er nøyaktig utskrift/kopi av ovennevnte søknad, som opprinnelig inngitt 2005.05.27

*It is hereby certified that the annexed document is a true copy of the above-mentioned application, as originally filed on 2005.05.27*

Priority is claimed from patent application no 20045612 filed on 2004.12.23

2006.01.04

Ellen B. Olsen
Saksbehandler

PATENTSTYRET
Styret for det industrielle rettsvern

# Søknad om **patent**

www.patentstyret.no

Ferdig utfylt skjema sendes til adressen nedenfor. Vennligst ikke heft sammen sidene.
Vi ber om at blankettene utfylles *maskinelt* eller ved bruk av *blokkbokstaver*. Skjema for utfylling på datamaskin kan lastes ned fra **www.patentstyret.no**.

2005 -05- 27

SØKNAD s. 1 av 2

## Søker – Den som søker om patent blir også innehaver av en eventuell rettighet. Må fylles ut.

Foretakets navn (fornavn hvis søker er person):
Jan O.

Etternavn (hvis søker er person):
Aasly

☐ Kryss av hvis søker tidligere har vært kunde hos Patentstyret.

Oppgi gjerne kundenummer:

Adresse:
Planetvegen 41

Postnummer:
7036

Poststed:
Trondheim

Land:

☒ Kryss av hvis flere søkere er angitt i medfølgende skjema eller på eget ark.

☒ Kryss av hvis søker(ne) utfører 20 årsverk eller mindre (se veiledning).

FLERE SØKERE

## Kontaktinfo – Hvem skal Patentstyret henvende seg til? Oppgi telefonnummer og eventuell referanse.

Fornavn til kontaktperson for fullmektig *eller* søker:
Bodil Merete

Etternavn:
Sollie

☎ Telefon:

Referanse (maks. 30 tegn):
aasly

FLERE OPPFINNERE

PRIORITETER

## Fullmektig – Hvis du ikke har oppnevnt en fullmektig, kan du gå til neste punkt.

Foretakets navn (fornavn hvis fullmektig er person):
CURO AS

Etternavn (hvis fullmektig er person):

☐ Kryss av hvis fullmektig tidligere har vært kunde hos Patentstyret.

Oppgi gjerne kundenummer:

Adresse:
Postboks 38

Postnummer:
7231

Poststed:
Lundamo

Land:

VEILEDNING

## Oppfinner – Oppfinneren skal alltid oppgis, selv om oppfinner og søker er samme person.

Oppfinnerens fornavn:
Jan O.

Etternavn:
Aasly

☐ Kryss av hvis oppfinner tidligere har vært kunde hos Patentstyret.

Oppgi gjerne kundenummer:

Adresse:
Planetvegen 41

Postnummer:
7036

Poststed:
Trondheim

Land:

☒ Kryss av hvis flere oppfinnere er angitt i medfølgende skjema eller på eget ark.

PS:101 - PDF - 2005.04.07


**Adresse**
▸ Postboks 8160 Dep.
Københavngaten 10
0033 Oslo

**Telefon**
▸ 22 38 73 00
**Telefaks**
▸ 22 38 73 01

**Bankgiro**
▸ 8276.01.00192
**Organisasjonsnr.**
▸ 971526157 MVA

PATENTSTYRET®
Styret for det industrielle rettsvern

... søknad om **patent**

**www.patentstyret.no**

SØKNAD s. 2 AV 2

## Tittel – Gi en kort benevnelse eller tittel for oppfinnelsen (ikke over 256 tegn, inkludert mellomrom)

Tittel:
Polynucleotide

## PCT – Fylles bare ut hvis denne søknaden er en videreføring av en tidligere innlevert internasjonal søknad (PCT)

| | Inngivelsesdato (åååå.mm.dd): | Søknadsnummer: |
|---|---|---|
| PCT-søknadens dato og nummer: | PCT | / |

## Prioritetskrav – Hvis du ikke har søkt om denne oppfinnelsen tidligere, i et annet land eller i Norge, kan du gå videre til neste punkt

**Prioritet kreves på grunnlag av tidligere innlevert søknad i Norge eller utlandet:**

| | Inngivelsesdato (åååå.mm.dd): | Landkode: | Søknadsnummer: |
|---|---|---|---|
| Opplysninger om tidligere søknad. Ved flere krav skal tidligste prioritet angis her: | 2004.12.23 | NO | 20045612 |

☐ Flere prioritetskrav er angitt i medfølgende skjema, eller på eget ark.

## Biologisk materiale – Fylles bare ut hvis oppfinnelsen omfatter biologisk materiale

**Søknaden omfatter biologisk materiale. Deponeringssted og nummer må oppgis:**

Deponeringssted og nummer (benytt gjerne eget ark):

☐ Prøve av materiale skal bare utleveres til en særlig sakkyndig.

## Avdelt/utskilt – Hvis du ikke har søkt om patent i Norge tidligere, kan du gå videre til neste punkt

**Søknaden er avdelt eller utskilt fra tidligere levert søknad i Norge:**

| | | Dato (åååå.mm.dd): | Søknadsnummer: |
|---|---|---|---|
| ☐ Avdelt søknad | Informasjon om opprinnelig søknad/innsendt tilleggsmateriale | | |
| ☐ Utskilt søknad | | | |

## Annet

☐ Søknaden er også levert per telefaks. Oppgi dato (åååå.mm.dd):

☐ Jeg har fått utført forundersøkelse. Oppgi nr (årstall - nummer - bokstav):

## Vedlegg – Angi hvilken dokumentasjon av oppfinnelsen du legger ved, samt andre vedlegg

☒ Tegninger — Oppgi antall tegninger: 6

☒ Beskrivelse av oppfinnelsen

☒ Patentkrav

☐ Sammendrag på norsk

☐ Dokumentasjon av eventuelle prioritetskrav (prioritetsbevis)

☐ Oversettelse av internasjonal søknad (kun hvis PCT-felt over er fylt ut)

☒ Fullmaktsdokument(er)

☐ Overdragelsesdokument(er)

☐ Erklæring om retten til oppfinnelsen

☒ Annet: sekvenslister, også på diskett

## Dato/underskrift – Sjekk at du har fylt ut punktene under «Søker», «Oppfinner» og «Vedlegg». Signer søknaden

Sted og dato (blokkbokstaver):
Lundamo 26. mai 2005

Signatur:
Bodil Merete Sollie

Navn i blokkbokstaver:
Bodil Merete Sollie

**NB!** Søknadsavgiften vil bli fakturert for alle søknader (dvs. at søknadsavgiften ikke skal følge søknaden). Betalingsfrist er ca. 1 måned, se faktura.

PATENTSTYRET®
Styret for det industrielle rettsvern

PS-101 - PDF - 2004-04-07

2005 -05- 27

Vedleggsskjema:

# Flere søkere

www.patentstyret.no

Dette skjemaet benyttes som vedlegg til patentsøknaden for å oppgi flere søkere. **NB! Gi hver søker et nummer.** Personen oppgitt på søknadsskjemaet vil alltid bli registrert som nr. 01. Første angivelse på dette skjema vil være søker 02. Skjema for utfylling på datamaskin kan lastes ned fra **www.patentstyret.no.**

**Referanse** – Gjenta referansen fra kontaktinfo, eventuelt søkerens navn, som angitt på søknadsskjemaets førsteside. Må fylles ut.

Referanse:
aasly

**Søker nr.** 2

Fornavn og mellomnavn:
Zbigniew K.

Etternavn:
Wszolek

☐ Søker har tidligere vært kunde hos Patentstyret.

Oppgi gjerne kundenummer:

Adresse:
3868 Biggin Church Rd W.

Jacksonville, FL 32224

Postnummer: Poststed: Land:
USA

FLERE SØKERE

**Søker nr.** 3

Fornavn og mellomnavn:
Matthew J.

Etternavn:
Farrer

☐ Søker har tidligere vært kunde hos Patentstyret.

Oppgi gjerne kundenummer:

Adresse:
1858 Arden Way.

Jacksonville, FL 32250

Postnummer: Poststed: Land:
USA

**Søker nr.**

Fornavn og mellomnavn:

Etternavn:

☐ Søker har tidligere vært kunde hos Patentstyret.

Oppgi gjerne kundenummer:

Adresse:

Postnummer: Poststed: Land:

**Søker nr.**

Fornavn og mellomnavn:

Etternavn:

☐ Søker har tidligere vært kunde hos Patentstyret.

Oppgi gjerne kundenummer:

Adresse:

Postnummer: Poststed: Land:

PS-101 - PDF - 2004.04.07

**NB!** Ved behov for mer plass benyttes flere skjema eller eget ark.

PATENTSTYRET®
Styret for det industrielle rettsvern

Vedleggsskjema 2005 -05- 27

# Flere oppfinnere

www.patentstyret.no

**Dette skjemaet benyttes som vedlegg til patentsøknaden for å oppgi flere oppfinnere. NB! Gi hver oppfinner et nummer.** Personen oppgitt på søknadsskjemaet vil alltid bli registrert som nr. 01. Første angivelse på dette skjema vil være oppfinner 02. Skjema for utfylling på datamaskin kan lastes ned fra **www.patentstyret.no.**

**Referanse** Gjenta referansen fra kontaktinfo. Eventuelt søkerens navn som angitt på søknadsskjemaets første side. Maskinutfylles.

Referanse:
aasly

**Oppfinner nr.** 2

Fornavn og mellomnavn: Zbigniew K.
Etternavn: Wszolek

☐ Oppfinner har tidligere vært kunde hos Patentstyret. Oppgi gjerne kundenummer:

Adresse:
3868 Biggin Church Rd W.

Jacksonville, FL 32224

Postnummer: Poststed: Land: USA

**Oppfinner nr.** 3

Fornavn og mellomnavn: Matthew J.
Etternavn: Farrer

☐ Oppfinner har tidligere vært kunde hos Patentstyret. Oppgi gjerne kundenummer:

Adresse:
1858 Arden Way

Jacksonville, FL 32250

Postnummer: Poststed: Land: USA

FLERE OPPFINNERE

**Oppfinner nr.**

Fornavn og mellomnavn: Etternavn:

☐ Oppfinner har tidligere vært kunde hos Patentstyret. Oppgi gjerne kundenummer:

Adresse:

Postnummer: Poststed: Land:

**Oppfinner nr.**

Fornavn og mellomnavn: Etternavn:

☐ Oppfinner har tidligere vært kunde hos Patentstyret. Oppgi gjerne kundenummer:

Adresse:

Postnummer: Poststed: Land:

PS-101 - PDF - 2004-04-07

**NB!** Ved behov for mer plass benyttes flere skjema eller eget ark.

PATENTSTYRET®
Styret for det industrielle rettsvern

2005 -05- 27

Applicants and inventors: Aasly, Jan O., Planetvegen 41, 7036 Trondheim, Norway

Wszolek, Zbigniew K., 3868 Biggin Church Rd W., Jacksonville, FL 32224, USA

Farrer, Matthew J., 1858 Arden Way, Jacksoville, FL 32250, USA

Contact person: Aasly

Attorney: CURO AS, Postboks 38, 7231 Lundamo

Title: Polynucleotide

Present invention relates to a novel polynucleotide involved in heritable Parkinson's disease (PD), a novel polypeptide encoded by the polynucleotide, and a method for diagnosing heritable Parkinson's disease (PD).

**Background**

Parkinsonism (MIM168600) is a clinical syndrome characterized by bradykinesia, resting tremor, muscle rigidity, and postural instability (Gelb et al. 1999). The most common cause of parkinsonism is Parkinson's disease (PD). Second to Alzheimer's disease, PD is the most common neurodegenerative disorder affecting >1% of the population over 55 years of age (de Rijk et al. 1995). Neuropathological findings in PD are loss of pigmented neurons in the brainstem, *substantia nigra* and *locus ceruleus*, with intracellular Lewy body inclusions found within surviving neurons (Forno 1996).

Although PD is considered a sporadic disease, various hereditary forms of parkinsonism have been recognized (Vila and Przedborski 2004). A major breakthrough in recent years has been the mapping and cloning of a number of genes causing monogenic forms of parkinsonism. Genomic multiplication and missense mutations in the α-synuclein gene were initially identified in a small number of families with autosomal dominant parkinsonism (PARK1/4 [MIM 168601]) (Polymeropoulos et al. 1997; Kruger et al. 1998; Singleton et al. 2003; Chartier-Harlin et al. 2004; Farrer et al. 2004; Zarranz et al. 2004). Subsequently, α-synuclein antibodies were found to robustly stain Lewy bodies and Lewy neurites in the *substantia nigra* in familial and sporadic PD (Spillantini et al. 1997) and common genetic variability in the α-synuclein promoter has been implicated in sporadic PD (Pals et al. 2004).

Autosomal recessive mutations in three genes, *parkin*, *DJ-1* and *PINK1* have been linked with early-onset parkinsonism (<45 years at onset) (PARK2, PARK6 & PARK7 [MIM 602533, 602544 & 608309])(Kitada et al. 1998; Bonifati et al. 2003; Valente et al. 2004). A large number of pathogenic mutations and rearrangements have been identified in the *parkin* gene reviewed by (Mata et al. 2004), but mutations in DJ-1 and PINK-1 are rare (unpublished data).

Very recently, five pathogenic mutations were identified in a gene, leucine-rich repeat kinase 2 (*LRRK2*) in six families with autosomal-dominant parkinsonism, linked to the PARK8 locus [MIM 607060]) (Zimprich et al. 2004a). Paisan-Ruiz and colleagues independently confirmed these findings of two pathogenic mutations in a British and Basque families (Paisan-Ruiz et al. x2004).

**Object**

The object of the invention is to isolate a gene or polynucleotide proving inheritable parkinsonism, and to use presence of this gene to diagnose a patient before he/her gets sick. A further object is to use this gene or polynucleotide to transfect a microorganism or experimental animal in order to develop a new medicine for treating parkinsonism, or for preventing the onset of the illness.

**The invention**

Inheritable parkinsonism may be proved by the method according to the characterizing part of claim 5, and the other objects are met by a polynucleotid according to the characterizing part of claim 1, a recombinant vector according to claim 3, a DNA probe and a DNA primer according to claims 4 and 6 respectively, and a peptide according to claim 9.

The inventors have isolated a novel LRRK2 mutation, and this mutation may cause development of dominantly inherited PD. By screening healthy persons, one can state whether the healthy persons have the mutation, and thus most likely will develop the illness.

As the sequence of the mutated gene is known, microorganisms and further experimental animals may be transfected, in order to investigate for a new medicine to treat or prevent the onset of the illness.

In the following the invention will be described by reference to a study of PD patients and their families. Parts of the study are shown in figures, wherein

Figure 1 shows a schematic drawing of LRRK2 with predicted protein domains,

Figure 2 shows pedigrees of families with LRRK2 G2019S,

Figure 3 shows chromosome 12q12 STR markers on the disease haplotype (PARK 8),

Figure 4 shows probability of becoming affected by parkinsonism, in LRRK2 G2019S carriers, as a function of age, and

Figure 5 shows aligned amino acid sequences of the activation loop of different human kinases.

The inventors identified seven unrelated persons all having the new mutation, from 248 multiplex kindreds with dominantly inherited PD, and six further unrelated persons from three populaton-based series of persons with dominantly inhereted PD. These 13 persons

and their families made basis for the inventors' further work. Segregation and linkage analysis provides evidence for pathogenicity and an estimate of age-associated penetrance; haplotype analysis demonstrates the mutation originates from a common and ancient founder.

**Subjects and Methods**

*Study subjects*

The patients and controls were examined by neurologists specialized in movement disorders. A full history, including family history and neurological examination, was completed on each patient. Clinical diagnosis of PD required the presence of at least two of three cardinal signs (resting tremor, bradykinesia and rigidity), improvement from adequate dopaminergic therapy and the absence of atypical features or other causes of parkinsonism.

*LRRK2 sequencing and mutation screening*

Blood samples were taken and genomic DNA was extracted using standard techniques. Six families (families 194, 281, 3081, 3082, 3083 and 3211) were known to have a positive LOD-score for STR (Short Tandem Repeat) markers in the PARK8 locus (Zimprich et al. 2004b). Amplification of all 51 exons of the *LRRK2* gene was performed by polymerase chain reaction (PCR) in one patient having PD, from each of these six families. All PCRs were carried out for each primer set with 20-50 ng of template DNA in a total volume of 25µl using a final reaction concentration of 200 µM dNTP, 1x PCR-Buffer (Qiagen), 1x Q-Solution (Qiagen), and 0.8 µM of each primer. One unit of Taq polymerase (Qiagen) was added to each reaction. Amplification was performed using a 57-52°C-touchdown protocol over 38 cycles. The primers used for PCR amplification of *LRRK2* exons and for sequencing are available on request.

The nucleotide sequences of all PCR products were determined by direct sequencing. Each PCR product was cleaned by using a Millipore PCR purification plate. Three microliters of purified PCR product was used per sequencing reaction with 1 µl of either the forward or reverse PCR primer and 1 µl of BigDye reaction mix (Applied Biosystems). Electrophoresis was performed under standard conditions on an ABI 3730 automated sequencer (Applied Biosystems). All sequences were obtained with both forward and reverse primers. Sequences were analyzed with SeqScape software version 2.1.1 (Applied

Biosystems) and compared with published sequence of *LRRK2* (GenBank accession no. AY792511).

After identification of a heterozygous G2019S (G6055A) mutation in the proband of family 3215 (referred to as family 3211 in Zimprich et al, 2004b), we designed a probe employing TaqMan chemistry on an ABI7900 (Applied Biosystems) to screen for this mutation. First we examined 248 PD patients from families with a known family history, consistent with autosomal dominant transmission of a suspected causative gene. Then 377 Norwegian, 271 Irish and 100 Polish PD patients (constituting the three population series) were checked using this assay; and 2260 samples of healthy persons from similar populations were also included (1200 US American, 550 Norwegian, 330 Irish and 180 Polish subjects), the latter to be used as control samples. Mutations were confirmed by direct sequencing of PCR products from *LRRK2* exon 41. Finally, all participating family members of *LRRK2* G2019 mutation carriers (affected and unaffected) were screened for the mutation.

By 6055 G>A or G6055A it is meant that nucleotide number 6055 of the LRRK2 gene, counted from the 5'end of the polynucleotide, has changed from G (guanine) to A (adenine). This change also causes a change in the polypeptide encoded by the polynucleotide, and G2019S denotes a polynucleotide where aminoacid number 2019 is changed from G(Glycine) to S(Serine). These shortenings are wellknown to persons skilled of the art.

*Genotyping of STR markers*

Fourteen STR markers were genotyped in mutation carriers and all available family members, in all 13 families, for linkage analyses and to determine whether there was a particular haplotype associated with the *LRRK2* mutation. STR markers were chosen to span the PARK8 region including D12S87, D12S1648, D12S2080, D12S2194, D12S1048, D12S1301 and D12S1701. *LRRK2* is located between D12S2194 and D12S1048. We also developed seven novel STR markers in this region (shown in table 1 below) by searching for repeat polymorphisms using RepeatMasker of *in silico* BAC sequence (UCSC Human Genome Browser Web site). The labeling of these novel markers reflects their physical position relative to the start codon of *LRRK2*.

| Table 1. Novel chromosome 12 STR markers | | |
|---|---|---|
| Marker name | Primer sequence | Physical position (bp) On chromosome 12 |
| D12S2514 | F: 5'-TTGCAGCTGTAAGGAATTTGGG-3' | 38873779 |
| | R: 5'-GCATTCTTCAGCCTGAGACCC-3' | |
| D12S2515 | F: 5'-TGAAGGACACTGAACAAGATGG-3' | 38974140 |
| | R: 5'-GCCATAGTCCTTCCATAGTTCC-3' | |
| D12S2516 | F: 5'-CGCAGCGAGCATTGTACC-3' | 38989214 |
| | R: 5'-CTCGGAAAGTTTCCCAATTC-3' | |
| D12S2518 | F: 5'-CTGGTATTACCTCAACTGTGGCTC-3' | 39034800 |
| | R: 5'-ACTGGTATGTTTAAGCCTGGCAC-3' | |
| D12S2519 | F: 5'-AGCAGCAGAGAAGATTTCAATAAC-3' | 39116816 |
| | R: 5'-AATCATCTTTGAAAGAACCAGG-3' | |
| D12S2523 | F: 5'-TAAACGAAGCTCCCTCACTGTAAG-3' | 39147728 |
| | R: 5'-TCTTTGTAGCTGCGGTTGTTTC-3' | |
| D12S2517 | F: 5'-TCATGAAGATGTCTGTGATAGGGC-3' | 39282976 |
| | R: 5'-CTCTATTGTGAGCAAACTGCATGG-3' | |

One primer of each pair was labeled with a fluorescent tag. PCR reactions were carried out on 10-20 ng of DNA in a total volume of 15 µl with final reaction concentrations of 150 µM dNTP, 1x PCR-Buffer (Qiagen), 1x Q-Solution (Qiagen) and 0.6 µM of each primer, with 1 unit of Taq Polymerase (Qiagen). Amplification was performed using a 57-52°C-touchdown protocol over 38 cycles. The PCR product for each marker was diluted by a factor of 10 to 100 with water. One microliter was then added to 10 µl of Hi-Di Formamide and Rox size standard. All samples were run on an ABI 3100 genetic analyzer, and results were analyzed using Genescan 3.7 and Genotyper 3.7 software (Applied Biosystems). Since population allele frequencies were not available from the CEPH database, these have been estimated by genotyping 95 unrelated Caucasian subjects, a population based series from the United States (shown in table 2 below).

| Table 2. Allele frequencies of Park 8 Markers | |
|---|---|
| **Marker and allele (bp)** | **Frequency (%)** |
| D12S87 (n = 92) | |
| 150 | 0.5 |
| 154 | 1.1 |
| 156 | 27.2 |
| 158 | 33.2 |
| 160 | 11.4 |
| 162 | 2.7 |
| 164 | 6.0 |
| 166 | 17.4 |
| 168 | 0.5 |
| D12S1648 (n = 91) | |
| 110 | 13.7 |
| 112 | 3.3 |
| 114 | 11.0 |
| 116 | 4.4 |
| 118 | 2.2 |
| 120 | 2.8 |
| 122 | 17.0 |
| 124 | 3.9 |
| 126 | 7.7 |
| 128 | 14.3 |
| 130 | 8.8 |
| 132 | 2.8 |
| 134 | 2.8 |
| 136 | 1.7 |
| 138 | 0.6 |
| 140 | 2.2 |
| 142 | 1.1 |
| D12S2080 (n = 93) | |
| 176 | 1.6 |
| 180 | 20.2 |
| 184 | 44.7 |
| 188 | 22.9 |
| 192 | 10.6 |
| D12S2194 (n = 87) | |
| 245 | 0.6 |
| 249 | 40.9 |
| 253 | 32.4 |
| 257 | 19.9 |
| 261 | 4.6 |
| 265 | 1.7 |
| D12S2514 (n = 82) | |
| 284 | 11.0 |
| 291 | 53.1 |
| 294 | 32.3 |
| 297 | 1.2 |
| 300 | 2.4 |

| Table 2. Allele frequencies of Park 8 Markers | |
|---|---|
| **Marker and allele (bp)** | **Frequency (%)** |
| D12S2515 (n = 93) | |
| 208 | 3.2 |
| 212 | 26.6 |
| 216 | 18.6 |
| 220 | 22.9 |
| 224 | 20.7 |
| 228 | 5.3 |
| 232 | 2.7 |
| rs 7966550 (n = 90) | |
| T | 90.6 |
| C | 9.4 |
| DS12S2516 | |
| 252 | 37.3 |
| 254 | 62.7 |
| rs 1427263 (n = 89) | |
| A | 63.6 |
| C | 36.5 |
| rs1116013 (n = 88) | |
| A | 49.4 |
| G | 50.6 |
| rs11564148 (n = 88) | |
| A | 26.1 |
| T | 73.9 |
| D12S2518 (N = 90) | |
| 154 | 79.7 |
| 168 | 15.9 |
| 170 | 4.4 |
| D12S519 (n = 72) | |
| 132 | 29.5 |
| 134 | 22.6 |
| 138 | 22.6 |
| 140 | 25.3 |
| D12S2520 (N =85) | |
| 248 | 8.2 |
| 251 | 7.6 |
| 254 | 10.0 |
| 257 | 54.1 |
| 260 | 20.0 |

| Table 2. Allele frequencies of Park 8 Markers | |
|---|---|
| **Marker and allele (bp)** | **Frequency (%)** |
| D12S2521 (N = 93) | |
| 311 | 0.5 |
| 315 | 10.8 |
| 319 | 20.4 |
| 323 | 8.1 |
| 327 | 7.0 |
| 331 | 8.1 |
| 335 | 0.5 |
| 355 | 1.1 |
| 359 | 7.5 |
| 363 | 13.4 |
| 367 | 7.0 |
| 371 | 7.0 |
| 375 | 6.5 |
| 379 | 3.8 |
| 383 | 1.1 |
| 387 | .5 |
| D12S2522 (N = 93) | |
| 281 | 9.1 |
| 283 | 14.0 |
| 285 | .5 |
| 287 | 11.3 |
| 293 | .5 |
| 295 | 15.6 |
| 297 | 44.6 |
| 299 | 4.3 |
| D12S2523 (n = 89) | |
| 305 | 18.9 |
| 314 | 41.1 |
| 317 | 8.9 |
| 320 | 30.0 |
| 323 | 1.1 |

**Table 2. Allele frequencies of Park 8 Markers**

| Marker and allele (bp) | Frequency (%) |
|---|---|
| D12S2517 (n = 93) | |
| 180 | 8.5 |
| 182 | 7.5 |
| 184 | 15.4 |
| 186 | 8.5 |
| 188 | 11.7 |
| 190 | 8.0 |
| 192 | 5.3 |
| 194 | 1.1 |
| 196 | 1.1 |
| 198 | 3.2 |
| 200 | 0.5 |
| 202 | 3.7 |
| 204 | 6.9 |
| 206 | 6.9 |
| 208 | 4.3 |
| 210 | 2.1 |
| 212 | 3.2 |
| 214 | 1.6 |
| 216 | 0.5 |
| D12S1048 (n = 89) | |
| 211 | 37.2 |
| 214 | 21.1 |
| 217 | 17.8 |
| 220 | 2.2 |
| 223 | 6.7 |
| 226 | 11.7 |
| 229 | 3.3 |
| D12S1301 (n = 93) | |
| 96 | 0.5 |
| 100 | 37.2 |
| 104 | 17.6 |
| 108 | 11.1 |
| 112 | 12.2 |
| 116 | 13.3 |
| 120 | 7.5 |
| 124 | 0.5 |

| Table 2. Allele frequencies of Park 8 Markers | |
|---|---|
| Marker and allele (bp) | Frequency (%) |
| D12S1701 (n = 93) | |
| 89 | 4.3 |
| 91 | 4.8 |
| 93 | 10.8 |
| 95 | 40.0 |
| 97 | 16.0 |
| 99 | 12.4 |
| 101 | 11.8 |
| 103 | 0.5 |

A The number of individuals genotyped is given for each marker (n)
B Alle frequencies are for individual markers in U.S. control subjects

*Statistical Analysis*

Multipoint nonparametric LOD scores for all families were calculated using GENEHUNTER-PLUS (Kong and Cox 1997). The frequency of the deleterious allele was set at 0.0001, and empirically determined allele frequencies were employed. The map positions for each marker were taken from Rutgers combined linkage-physical map version 1.0 (MAP-O-MAT web site). The three loci D12S2080, D12S2194 and D12S1301 are very tightly linked, with no observed recombinants in the database or within our genotyped families, and thus inter-marker distances were assigned as 0.01cM.

Chromosome 12 haplotypes in the PARK8 region were established for those families in which chromosome phase for mutation-carrying individuals could be deduced, thereby determining which alleles co-segregated with the *LRRK2* G2019S mutation in each family. For those affected individuals in whom the associated allele for a marker could not be determined, both alleles are given.

The age-dependent penetrance was estimated as the probability of a gene carrier becoming affected, at a given age, within the 13 families. The number of affected mutation carriers, for each decade, was divided by the total number of affected individuals, plus the number of unaffected carriers within that range. For some affected family members no DNA was available and only historical data on the disease course was obtained. These individuals were excluded from penetrance calculations.

**Results**

As mentioned previously, we identified 13 affected probands (i.e. 13 patients) who carry a heterozygous G6055A mutation in exon 41 of the *LRRK2* gene. The mutation leads

to a G2019S amino acid substitution of a highly conserved residue within the predicted activation loop of the MAPKKK (Mitogen-Activated Protein Kinase Kinase Kinase) domain (figure 1). After genotyping a total of 42 additional family members, 22 additional subjects were found to carry the mutation, seven with a diagnosis of PD (shown in table 3 below). One affected member of family P-089 did not carry the mutation and, for the purposes of this study, was considered a phenocopy and excluded from further analyses. Seven families originated from Norway, three were from the United States, two from Ireland, and one was from Poland. One family from the United States descended from Russian/Rumania, and another from Italy. For only one family (family 111), the ethnic origin was unknown. The *LRRK2* G2019S mutation segregates with disease in all kindreds, consistent with autosomal dominant transmission. To ensure patient confidentiality, simplified versions of the family pedigrees are presented in figure 2. There was no evidence of the mutation in the 2260 control samples.

Age at onset of clinical symptoms was quite variable, even within the same family. Family 1120, a family from the United States, had both the earliest and latest age at onset for a patient. The youngest affected subject had an onset at 39 years, whereas the oldest carrier presented with initial symptoms at 78 years. Where recorded, most *LRRK2* G2019S carriers have late-onset disease (>50 years at onset). The mean age at onset of affected mutation carriers was 56.8 years (range 39-78 years, n=19). Unaffected carriers have a mean age of 53.9 years (range 26-74 years, n=14). The penetrance of the mutation was found to be highly age-dependent, increasing from 17% at the age of 50 to 85% at the age of 70 (figure 4).

**Table 3**

**Demographic and Clinical Information for 13 Families with *LRRK2* G2019S**

| CHARACTERISTIC | FINDINGS FOR FAMILY | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Country of origin | P-063 Norway | P-089 Norway | P-104 Norway | P-241 Norway | P-369 Norway | P-394 Norway | F05 Norway | I210 United States | I11 United States | I120 United States | PD66 Ireland | 3211 Ireland | IP Poland |
| No. of generations | 3 | 4 | 3 | 3 | 3 | 4 | 4 | 2 | 2 | 3 | 1 | 2 | 1 |
| No. of affected individuals | 2 | 4 | 4 | 1 | 3 | 4 | 5 | 2 | 3 | 3 | -1 | 3 | 1 |
| No. of typed individuals affected (unaffected) | 1 (6) | 2 (8) | 1 (1) | 1 (4) | 2 (3) | 1 (1) | 3 (6) | 1 (0) | 2(0) | 3 (3) | 1 (0) | 2 (6) | 1 (0) |
| No. of typed generations | 2 | 3 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 1 |
| Age[a] at onset in years (range) | 59 (53–65) | 59 (43–70) | 58 | 60 | 50 (43–61) | 66 | 64 (61–70) | 65 | 58 (57–58) | 59 (39–78) | 41 | 46 (40–52) | 73 |
| Maximum mLOD score | 0 | .30 | 0 | 0 | .60 | 0 | .90 | 0 | .09 | .30 | 0 | .30 | 0 |

[a]Average ages at onset are given when affected individuals. $n \geq 2$

Evidence for linkage (the statistical burden of proof that this mutation causes disease) to the PARK8 locus was found across families, with a combined maximum multipoint LOD score of 2.41 [for all 14 markers], corresponding to a P value of $4{,}3 \times 10^{-4}$. As only a defined chromosomal region was investigated, rather than a genome-wide search, this LOD score exceeds that required for significance, P=0.01 (Lander and Kruglyak 1995). A positive LOD score was found in all families where more then one affected subject was genotyped (table 3).

All affected members from the different families, except the individual in family P-089 who did not carry the mutation, appear to share a common haplotype on chromosome 12 the *LRRK2* gene locus (figure 3). Haplotypes can be established with certainty in nine of the families, and all mutation carriers in these families share alleles for four STR markers and 4 single nucleotide polymorphisms (SNPs) in the *LRRK2* gene locus. These markers are *LRRK2* D12S2516, D12S2518, D12S2519, D12S2520 and SNPs rs7966550, rs1427263, rs11176013, rs11564148. For the remaining families, the number of available samples from relatives was not sufficient to determine phase. However, the genotypes in these cases are consistent with a common *LRRK2* G2019S allele. D12S2516 is located in intron 29 and D12S2518 is located in intron 44 of the *LRRK2* gene, whereas the two other shared markers are positioned 3' of the gene. Using the physical position of the shared and non-shared markers, the size of the shared haplotype is between 145 kb and 154 kb.

**Discussion**

We have identified a novel *LRRK2* mutation, G2019S, which co-segregates with autosomal dominant parkinsonism in 13 kindreds originating from several European populations. Positive LOD scores were obtained in multiplex families, and combined they provide significant support for the PARK8 locus. *LRRK2* G2019S mutation was absent in a large number of control subjects, and of similar ethnicity. The number of families linked to *LRRK2* in this and previous studies now explains the majority of genetically defined autosomal dominant parkinsonism.

The mean age at onset of affected *LRRK2* G2019S carriers was 56.8 years, and comparable to that of patients in other families linked to PARK8 (Funayama et al. 2002; Paisan-Ruiz et al. 2004; Zimprich et al. 2004a). The majority of patients present with late-onset disease, indistinguishable from typical idiopathic PD. Disease penetrance is age-dependent, and increases in a linear fashion from 17% at the age of 50 to 85% at the age of 70. Age is the single most consistent risk factor for development of PD and other neurodegenerative disorders (Lang and Lozano 1998), and an important risk factor in *LRRK2* associated parkinsonism. Interestingly, age at onset was variable in this study, both within and between different families, suggesting other susceptibility factors, environmental or genetic, may influence the phenotype.

Although our findings clearly indicate that *LRRK2* mutations account for a substantial proportion of familial late-onset parkinsonism, historically, cross-sectional twin studies have not supported a genetic etiology for late-onset PD (Tanner et al. 1999; Wirdefeldt et

al. 2004). The age-associated penetrance of *LRRK2* mutations provides some explanation as even large and well designed twin studies are underpowered to detect incompletely penetrant mutations (Simon et al. 2002). *LRRK2* mutations were also found in apparently sporadic PD patients; three of the patients in this study did not have any known affected first- or second-degree relatives. However, a caveat of age-dependent penetrance is that carriers may die of other diseases, before manifesting or being diagnosed with PD. Thus, it seems difficult to separate sporadic and familial PD, or to hypothesize environmental causes to be more important in one group and genetic causes more prominent in the other. In light of these results, a family history of parkinsonism, previously considered an exclusion criterion for a diagnosis of PD, must be reconsidered (Hughes et al. 1992).

*LRRK2* is a member of the recently defined ROCO protein family (Bosgraaf and Van Haastert 2003). In human, mouse and rat, members of the ROCO protein family have five conserved domains (figure 1). The kinase domain belongs to the MAPKKK subfamily of kinases. The active sites of all kinases are located in a cleft between an N-terminal and a C-terminal lobe, typically covered by an 'activation loop', in an inactive conformation. The activation loop must undergo crucial structural changes to allow access to peptide substrates and to orientate key catalytic amino acids (Huse and Kuriyan 2002). In different kinases, the activation loop starts and ends with the conserved residues asp-phe-gly (DFG) and ala-pro-glu (APE), respectively (Dibb et al. 2004). Of note, the *LRRK2* G2019S substitution changes a highly conserved amino acid at the start of this loop (figure 5). In a German family we previously described, an I2020T mutation is located in an adjacent codon (Zimprich et al. 2004a). In other kinases, oncogenic mutations in residues within the activation loop of the kinase domain have an activating effect (Davies et al. 2002), thus we postulate *LRRK2* G2019S and I2020T mutations may have an effect on its kinase activity.

The age of an allele may be estimated from the genetic variation among different copies (intra-allelic variation), or from its frequency (Slatkin and Rannala 2000). However, the local recombination rate on chromosome 12q12 is unknown, as is the frequency of the G2019S mutation in the general population. Nevertheless, at centromeres there is generally a dearth in recombination; indeed no crossovers have been observed between *LRRK2* flanking markers D12S2194 and D12S1048 in our studies, or within CEPH families (MAP-O-MAT web site). The physical size of the shared haplotype is also small, between 145 kb and 154 kb, and the allele is widespread in families from several European populations. Hence, the mutation is likely to be ancient and may be relatively common in specific

populations. These data suggest a substantial proportion of late-onset PD will have a genetic basis.

**Electronic-Database Information**

The physical position of markers is from NCBI build 34. Accession numbers and URLs for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/

MAP-O-MAT, http://compgen.rutgers.edu/mapomat

RepeatMasker, http://www.repeatmasker.org/

**References**

Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, Dekker MC, Squitieri F, Ibanez P, Joosse M, van Dongen JW, Vanacore N, van Swieten JC, Brice A, Meco G, van Duijn CM, Oostra BA, Heutink P (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 299:256-9

Bosgraaf L, Van Haastert PJ (2003) Roc, a Ras/GTPase domain in complex proteins. Biochim Biophys Acta 1643:5-10

Chartier-Harlin MC, Kachergus J, Roumier C, Mouroux V, Douay X, Lincoln S, Levecque C, Larvor L, Andrieux J, Hulihan M, Waucquier N, Defebvre L, Amouyel P, Farrer M, Destee A (2004) Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. Lancet 364:1167-9

Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, et al. (2002) Mutations of the BRAF gene in human cancer. Nature 417:949-54

de Rijk MC, Breteler MM, Graveland GA, Ott A, Grobbee DE, van der Meche FG, Hofman A (1995) Prevalence of Parkinson's disease in the elderly: the Rotterdam Study. Neurology 45:2143-6

Dibb NJ, Dilworth SM, Mol CD (2004) Switching on kinases: oncogenic activation of BRAF and the PDGFR family. Nat Rev Cancer 4:718-27

Farrer M, Kachergus J, Forno L, Lincoln S, Wang DS, Hulihan M, Maraganore D, Gwinn-Hardy K, Wszolek Z, Dickson D, Langston JW (2004) Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. Ann Neurol 55:174-9

Forno LS (1996) Neuropathology of Parkinson's disease. J Neuropathol Exp Neurol 55:259-72

Funayama M, Hasegawa K, Kowa H, Saito M, Tsuji S, Obata F (2002) A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. Ann Neurol 51:296-301

Gelb DJ, Oliver E, Gilman S (1999) Diagnostic criteria for Parkinson disease. Arch Neurol 56:33-9

Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55:181-4

Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. Cell 109:275-82

Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392:605-8

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179-88

Kruger R, Kuhn W, Muller T, Woitalla D, Graeber M, Kosel S, Przuntek H, Epplen JT, Schols L, Riess O (1998) Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat Genet 18:106-8

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241-7

Lang AE, Lozano AM (1998) Parkinson's disease. First of two parts. N Engl J Med 339:1044-53

Mata IF, Lockhart PJ, Farrer MJ (2004) Parkin genetics: one model for Parkinson's disease. Hum Mol Genet 13 Spec No 1:R127-33

Paisan-Ruiz C, Jain S, Evans EW, Gilks WP, Simon J, van der Brug M, de Munain AL, Aparicio S, Gil AM, Khan N, Johnson J, Martinez JR, Nicholl D, Carrera IM, Pena AS, de Silva R, Lees A, Marti-Masso JF, Perez-Tur J, Wood NW, Singleton AB (2004) Cloning of the Gene Containing Mutations that Cause PARK8-Linked Parkinson's Disease. Neuron 44:595-600

Pals P, Lincoln S, Manning J, Heckman M, Skipper L, Hulihan M, Van den Broeck M, De Pooter T, Cras P, Crook J, Van Broeckhoven C, Farrer MJ (2004) alpha-Synuclein promoter confers susceptibility to Parkinson's disease. Ann Neurol 56:591-5

Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 276:2045-7

Simon DK, Lin MT, Pascual-Leone A (2002) "Nature versus nurture" and incompletely penetrant mutations. J Neurol Neurosurg Psychiatry 72:686-9

Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muenter M, Baptista M, Miller D, Blancato J, Hardy J, Gwinn-Hardy K (2003) alpha-Synuclein locus triplication causes Parkinson's disease. Science 302:841

Slatkin M, Rannala B (2000) Estimating allele age. Annu Rev Genomics Hum Genet 1:225-49

Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M (1997) Alpha-synuclein in Lewy bodies. Nature 388:839-40

Tanner CM, Ottman R, Goldman SM, Ellenberg J, Chan P, Mayeux R, Langston JW (1999) Parkinson disease in twins: an etiologic study. Jama 281:341-6

Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, Ali Z, Del Turco D, Bentivoglio AR, Healy DG, Albanese A, Nussbaum R, Gonzalez-Maldonado R, Deller T, Salvi S, Cortelli P, Gilks WP, Latchman DS, Harvey RJ, Dallapiccola B, Auburger G, Wood NW (2004) Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 304:1158-60

Vila M, Przedborski S (2004) Genetic clues to the pathogenesis of Parkinson's disease. Nat Med 10 Suppl:S58-62

Wirdefeldt K, Gatz M, Schalling M, Pedersen NL (2004) No evidence for heritability of Parkinson disease in Swedish twins. Neurology 63:305-11

Zarranz JJ, Alegre J, Gomez-Esteban JC, Lezcano E, Ros R, Ampuero I, Vidal L, Hoenicka J, Rodriguez O, Atares B, Llorens V, Gomez Tortosa E, del Ser T, Munoz DG, de Yebenes JG (2004) The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann Neurol 55:164-73

Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, Kachergus J, Hulihan M, Uitti RJ, Calne DB, Stoessl AJ, Pfeiffer RF, Patenge N, Carbajal IC, Vieregge P, Asmus F, Muller-Myhsok B, Dickson DW, Meitinger T, Strom TM, Wszolek ZK, Gasser T

(2004a) Mutations in LRRK2 Cause Autosomal-Dominant Parkinsonism with Pleomorphic Pathology. Neuron 44:601-7

Zimprich A, Muller-Myhsok B, Farrer M, Leitner P, Sharma M, Hulihan M, Lockhart P, Strongosky A, Kachergus J, Calne DB, Stoessl J, Uitti RJ, Pfeiffer RF, Trenkwalder C, Homann N, Ott E, Wenzel K, Asmus F, Hardy J, Wszolek Z, Gasser T (2004b) The PARK8 locus in autosomal dominant parkinsonism: confirmation of linkage and further delineation of the disease-containing interval. Am J Hum Genet 74:11-9

**Patent claims**

1. A polynucleotide consisting of the base sequence of SEQ ID NO: 2, or a complementary strand thereto, wherein the X is one of the group being defined by the bases A, C or T

2. A polynucleotide according to claim 1, wherein the polynucleotide is at the least a part of a gene.

3. A recombinant vector comprising a polynucleotid according to claim 1.

4. A DNA probe specific for the polynucleotide of claim 1, wherein it contains more than 10 consecutive nucleotides from the nucleotide, or the complementary strand.

5. A method of proving parkinsonism inheritance, by screening a sample of material taken from the subject of interest, with a probe according to claim 5.

6. DNA primer specific for the polynucleotide of claim 1, wherein it contains more than 10 consecutive nucleotides from the nucleotide, or the complementary strand.

7. Use of a polynucleotide according to claim 1, or a vector according to claim 4, to transfect an organism.

8. Use according to claim 7, wherein the organism is a mammal.

9. A peptide consisting of the base sequence of SEQ ID NO:1, wherein the x is not glycine.

**Abstract:**

A polynucleotide consisting of the base sequence of SEQ ID NO: 2, or a complementary strand thereto, wherein the X is one of the group being defined by the bases A, C or T. A primer and a probe specific for that polynucleotide, wherein the primer and/or probe contains at the least 10 consecutive nucleotides, and finally use of the probe for proving parkinsonism inheritance.

SEQUENCE LISTING

<110> Aasly, Jan O
Wszolek, Zbigniew K
Farrer, Matthew J

<120> Polynucleotide

<130> U1913

<150> NO20045612
<151> 2004-12-23

<160> 2

<170> PatentIn version 3.1

<210> 1
<211> 2527
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (2019)..(2019)
<223> X may be any amino acid except glycine.

<300>
<308> Genbank / AY792511
<309> 2004-11-15
<313> (1)..(2527)

<400> 1

Met Ala Ser Gly Ser Cys Gln Gly Cys Glu Glu Asp Glu Glu Thr Leu
1 5 10 15

Lys Lys Leu Ile Val Arg Leu Asn Asn Val Gln Glu Gly Lys Gln Ile
20 25 30

Glu Thr Leu Val Gln Ile Leu Glu Asp Leu Leu Val Phe Thr Tyr Ser
35 40 45

Glu His Ala Ser Lys Leu Phe Gln Gly Lys Asn Ile His Val Pro Leu
50 55 60

Leu Ile Val Leu Asp Ser Tyr Met Arg Val Ala Ser Val Gln Gln Val
65 70 75 80

Gly Trp Ser Leu Leu Cys Lys Leu Ile Glu Val Cys Pro Gly Thr Met
85 90 95

Gln Ser Leu Met Gly Pro Gln Asp Val Gly Asn Asp Trp Glu Val Leu
100 105 110

Gly Val His Gln Leu Ile Leu Lys Met Leu Thr Val His Asn Ala Ser
115 120 125

Val Asn Leu Ser Val Ile Gly Leu Lys Thr Leu Asp Leu Leu Leu Thr
130 135 140

Ser Gly Lys Ile Thr Leu Leu Ile Leu Asp Glu Glu Ser Asp Ile Phe
145 150 155 160

Met Leu Ile Phe Asp Ala Met His Ser Phe Pro Ala Asn Asp Glu Val
165 170 175

Gln Lys Leu Gly Cys Lys Ala Leu His Val Leu Phe Glu Arg Val Ser
180 185 190

Glu Glu Gln Leu Thr Glu Phe Val Glu Asn Lys Asp Tyr Met Ile Leu
195 200 205

Leu Ser Ala Ser Thr Asn Phe Lys Asp Glu Glu Glu Ile Val Leu His
210 215 220

Val Leu His Cys Leu His Ser Leu Ala Ile Pro Cys Asn Asn Val Glu
225 230 235 240

Val Leu Met Ser Gly Asn Val Arg Cys Tyr Asn Ile Val Val Glu Ala
245 250 255

Met Lys Ala Phe Pro Met Ser Glu Arg Ile Gln Glu Val Ser Cys Cys
260 265 270

Leu Leu His Arg Leu Thr Leu Gly Asn Phe Phe Asn Ile Leu Val Leu
275 280 285

Asn Glu Val His Glu Phe Val Val Lys Ala Val Gln Gln Tyr Pro Glu
290 295 300

Asn Ala Ala Leu Gln Ile Ser Ala Leu Ser Cys Leu Ala Leu Leu Thr
305 310 315 320

Glu Thr Ile Phe Leu Asn Gln Asp Leu Glu Glu Lys Asn Glu Asn Gln
325 330 335

Glu Asn Asp Asp Glu Gly Glu Glu Asp Lys Leu Phe Trp Leu Glu Ala
340 345 350

Cys Tyr Lys Ala Leu Thr Trp His Arg Lys Asn Lys His Val Gln Glu
355 360 365

Ala Ala Cys Trp Ala Leu Asn Asn Leu Leu Met Tyr Gln Asn Ser Leu
370 375 380

His Glu Lys Ile Gly Asp Glu Asp Gly His Phe Pro Ala His Arg Glu
385 390 395 400

Val Met Leu Ser Met Leu Met His Ser Ser Ser Lys Glu Val Phe Gln
405 410 415

Ala Ser Ala Asn Ala Leu Ser Thr Leu Leu Glu Gln Asn Val Asn Phe
420 425 430

Arg Lys Ile Leu Leu Ser Lys Gly Ile His Leu Asn Val Leu Glu Leu
435 440 445

Met Gln Lys His Ile His Ser Pro Glu Val Ala Glu Ser Gly Cys Lys
450 455 460

Met Leu Asn His Leu Phe Glu Gly Ser Asn Thr Ser Leu Asp Ile Met
465 470 475 480

Ala Ala Val Val Pro Lys Ile Leu Thr Val Met Lys Arg His Glu Thr
485 490 495

Ser Leu Pro Val Gln Leu Glu Ala Leu Arg Ala Ile Leu His Phe Ile
500 505 510

Val Pro Gly Met Pro Glu Glu Ser Arg Glu Asp Thr Glu Phe His His
515 520 525

Lys Leu Asn Met Val Lys Lys Gln Cys Phe Lys Asn Asp Ile His Lys
530 535 540

Leu Val Leu Ala Ala Leu Asn Arg Phe Ile Gly Asn Pro Gly Ile Gln
545 550 555 560

Lys Cys Gly Leu Lys Val Ile Ser Ser Ile Val His Phe Pro Asp Ala
565 570 575

Leu Glu Met Leu Ser Leu Glu Gly Ala Met Asp Ser Val Leu His Thr
580 585 590

Leu Gln Met Tyr Pro Asp Asp Gln Glu Ile Gln Cys Leu Gly Leu Ser
595 600 605

Leu Ile Gly Tyr Leu Ile Thr Lys Lys Asn Val Phe Ile Gly Thr Gly
610 615 620

His Leu Leu Ala Lys Ile Leu Val Ser Ser Leu Tyr Arg Phe Lys Asp
625 630 635 640

Val Ala Glu Ile Gln Thr Lys Gly Phe Gln Thr Ile Leu Ala Ile Leu
645 650 655

Lys Leu Ser Ala Ser Phe Ser Lys Leu Leu Val His His Ser Phe Asp
660 665 670

Leu Val Ile Phe His Gln Met Ser Ser Asn Ile Met Glu Gln Lys Asp
675 680 685

Gln Gln Phe Leu Asn Leu Cys Cys Lys Cys Phe Ala Lys Val Ala Met
690 695 700

Asp Asp Tyr Leu Lys Asn Val Met Leu Glu Arg Ala Cys Asp Gln Asn
705 710 715 720

Asn Ser Ile Met Val Glu Cys Leu Leu Leu Leu Gly Ala Asp Ala Asn
725 730 735

Gln Ala Lys Glu Gly Ser Ser Leu Ile Cys Gln Val Cys Glu Lys Glu
740 745 750

Ser Ser Pro Lys Leu Val Glu Leu Leu Leu Asn Ser Gly Ser Arg Glu
755 760 765

Gln Asp Val Arg Lys Ala Leu Thr Ile Ser Ile Gly Lys Gly Asp Ser
770 775 780

Gln Ile Ile Ser Leu Leu Leu Arg Arg Leu Ala Leu Asp Val Ala Asn
785 790 795 800

Asn Ser Ile Cys Leu Gly Gly Phe Cys Ile Gly Lys Val Glu Pro Ser
805 810 815

Trp Leu Gly Pro Leu Phe Pro Asp Lys Thr Ser Asn Leu Arg Lys Gln
820 825 830

Thr Asn Ile Ala Ser Thr Leu Ala Arg Met Val Ile Arg Tyr Gln Met

835 840 845

Lys Ser Ala Val Glu Glu Gly Thr Ala Ser Gly Ser Asp Gly Asn Phe
850 855 860

Ser Glu Asp Val Leu Ser Lys Phe Asp Glu Trp Thr Phe Ile Pro Asp
865 870 875 880

Ser Ser Met Asp Ser Val Phe Ala Gln Ser Asp Asp Leu Asp Ser Glu
885 890 895

Gly Ser Glu Gly Ser Phe Leu Val Lys Lys Lys Ser Asn Ser Ile Ser
900 905 910

Val Gly Glu Phe Tyr Arg Asp Ala Val Leu Gln Arg Cys Ser Pro Asn
915 920 925

Leu Gln Arg His Ser Asn Ser Leu Gly Pro Ile Phe Asp His Glu Asp
930 935 940

Leu Leu Lys Arg Lys Arg Lys Ile Leu Ser Ser Asp Asp Ser Leu Arg
945 950 955 960

Ser Ser Lys Leu Gln Ser His Met Arg His Ser Asp Ser Ile Ser Ser
965 970 975

Leu Ala Ser Glu Arg Glu Tyr Ile Thr Ser Leu Asp Leu Ser Ala Asn
980 985 990

Glu Leu Arg Asp Ile Asp Ala Leu Ser Gln Lys Cys Cys Ile Ser Val
995 1000 1005

His Leu Glu His Leu Glu Lys Leu Glu Leu His Gln Asn Ala Leu
1010 1015 1020

Thr Ser Phe Pro Gln Gln Leu Cys Glu Thr Leu Lys Ser Leu Thr
1025 1030 1035

His Leu Asp Leu His Ser Asn Lys Phe Thr Ser Phe Pro Ser Tyr
1040 1045 1050

Leu Leu Lys Met Ser Cys Ile Ala Asn Leu Asp Val Ser Arg Asn
1055 1060 1065

Asp Ile Gly Pro Ser Val Val Leu Asp Pro Thr Val Lys Cys Pro
1070 1075 1080

Thr Leu Lys Gln Phe Asn Leu Ser Tyr Asn Gln Leu Ser Phe Val
1085 1090 1095

Pro Glu Asn Leu Thr Asp Val Val Glu Lys Leu Glu Gln Leu Ile
1100 1105 1110

Leu Glu Gly Asn Lys Ile Ser Gly Ile Cys Ser Pro Leu Arg Leu
1115 1120 1125

Lys Glu Leu Lys Ile Leu Asn Leu Ser Lys Asn His Ile Ser Ser
1130 1135 1140

Leu Ser Glu Asn Phe Leu Glu Ala Cys Pro Lys Val Glu Ser Phe
1145 1150 1155

Ser Ala Arg Met Asn Phe Leu Ala Ala Met Pro Phe Leu Pro Pro
1160 1165 1170

Ser Met Thr Ile Leu Lys Leu Ser Gln Asn Lys Phe Ser Cys Ile
1175 1180 1185

Pro Glu Ala Ile Leu Asn Leu Pro His Leu Arg Ser Leu Asp Met
1190 1195 1200

Ser Ser Asn Asp Ile Gln Tyr Leu Pro Gly Pro Ala His Trp Lys
1205 1210 1215

Ser Leu Asn Leu Arg Glu Leu Leu Phe Ser His Asn Gln Ile Ser
1220 1225 1230

Ile Leu Asp Leu Ser Glu Lys Ala Tyr Leu Trp Ser Arg Val Glu
1235 1240 1245

Lys Leu His Leu Ser His Asn Lys Leu Lys Glu Ile Pro Pro Glu
1250 1255 1260

Ile Gly Cys Leu Glu Asn Leu Thr Ser Leu Asp Val Ser Tyr Asn
1265 1270 1275

Leu Glu Leu Arg Ser Phe Pro Asn Glu Met Gly Lys Leu Ser Lys
1280 1285 1290

Ile Trp Asp Leu Pro Leu Asp Glu Leu His Leu Asn Phe Asp Phe
1295 1300 1305

Lys His Ile Gly Cys Lys Ala Lys Asp Ile Ile Arg Phe Leu Gln
1310 1315 1320

Gln Arg Leu Lys Lys Ala Val Pro Tyr Asn Arg Met Lys Leu Met
1325 1330 1335

Ile Val Gly Asn Thr Gly Ser Gly Lys Thr Thr Leu Leu Gln Gln
1340 1345 1350

Leu Met Lys Thr Lys Lys Ser Asp Leu Gly Met Gln Ser Ala Thr
1355 1360 1365

Val Gly Ile Asp Val Lys Asp Trp Pro Ile Gln Ile Arg Asp Lys
1370 1375 1380

Arg Lys Arg Asp Leu Val Leu Asn Val Trp Asp Phe Ala Gly Arg
1385 1390 1395

Glu Glu Phe Tyr Ser Thr His Pro His Phe Met Thr Gln Arg Ala
1400 1405 1410

Leu Tyr Leu Ala Val Tyr Asp Leu Ser Lys Gly Gln Ala Glu Val
1415 1420 1425

Asp Ala Met Lys Pro Trp Leu Phe Asn Ile Lys Ala Arg Ala Ser
1430 1435 1440

Ser Ser Pro Val Ile Leu Val Gly Thr His Leu Asp Val Ser Asp
1445 1450 1455

Glu Lys Gln Arg Lys Ala Cys Met Ser Lys Ile Thr Lys Glu Leu
1460 1465 1470

Leu Asn Lys Arg Gly Phe Pro Ala Ile Arg Asp Tyr His Phe Val
1475 1480 1485

Asn Ala Thr Glu Glu Ser Asp Ala Leu Ala Lys Leu Arg Lys Thr
1490 1495 1500

Ile Ile Asn Glu Ser Leu Asn Phe Lys Ile Arg Asp Gln Leu Val
1505 1510 1515

Val Gly Gln Leu Ile Pro Asp Cys Tyr Val Glu Leu Glu Lys Ile
1520 1525 1530

Ile Leu Ser Glu Arg Lys Asn Val Pro Ile Glu Phe Pro Val Ile
1535 1540 1545

Asp Arg Lys Arg Leu Leu Gln Leu Val Arg Glu Asn Gln Leu Gln
1550 1555 1560

Leu Asp Glu Asn Glu Leu Pro His Ala Val His Phe Leu Asn Glu
1565 1570 1575

Ser Gly Val Leu Leu His Phe Gln Asp Pro Ala Leu Gln Leu Ser
1580 1585 1590

Asp Leu Tyr Phe Val Glu Pro Lys Trp Leu Cys Lys Ile Met Ala
1595 1600 1605

Gln Ile Leu Thr Val Lys Val Glu Gly Cys Pro Lys His Pro Lys
1610 1615 1620

Gly Ile Ile Ser Arg Arg Asp Val Glu Lys Phe Leu Ser Lys Lys
1625 1630 1635

Arg Lys Phe Pro Lys Asn Tyr Met Ser Gln Tyr Phe Lys Leu Leu
1640 1645 1650

Glu Lys Phe Gln Ile Ala Leu Pro Ile Gly Glu Glu Tyr Leu Leu
1655 1660 1665

Val Pro Ser Ser Leu Ser Asp His Arg Pro Val Ile Glu Leu Pro
1670 1675 1680

His Cys Glu Asn Ser Glu Ile Ile Ile Arg Leu Tyr Glu Met Pro
1685 1690 1695

Tyr Phe Pro Met Gly Phe Trp Ser Arg Leu Ile Asn Arg Leu Leu
1700 1705 1710

Glu Ile Ser Pro Tyr Met Leu Ser Gly Arg Glu Arg Ala Leu Arg
1715 1720 1725

Pro Asn Arg Met Tyr Trp Arg Gln Gly Ile Tyr Leu Asn Trp Ser
1730 1735 1740

Pro Glu Ala Tyr Cys Leu Val Gly Ser Glu Val Leu Asp Asn His
1745 1750 1755

Pro Glu Ser Phe Leu Lys Ile Thr Val Pro Ser Cys Arg Lys Gly

1760 1765 1770

Cys Ile Leu Leu Gly Gln Val Val Asp His Ile Asp Ser Leu Met
1775 1780 1785

Glu Glu Trp Phe Pro Gly Leu Leu Glu Ile Asp Ile Cys Gly Glu
1790 1795 1800

Gly Glu Thr Leu Leu Lys Lys Trp Ala Leu Tyr Ser Phe Asn Asp
1805 1810 1815

Gly Glu Glu His Gln Lys Ile Leu Leu Asp Asp Leu Met Lys Lys
1820 1825 1830

Ala Glu Glu Gly Asp Leu Leu Val Asn Pro Asp Gln Pro Arg Leu
1835 1840 1845

Thr Ile Pro Ile Ser Gln Ile Ala Pro Asp Leu Ile Leu Ala Asp
1850 1855 1860

Leu Pro Arg Asn Ile Met Leu Asn Asn Asp Glu Leu Glu Phe Glu
1865 1870 1875

Gln Ala Pro Glu Phe Leu Leu Gly Asp Gly Ser Phe Gly Ser Val
1880 1885 1890

Tyr Arg Ala Ala Tyr Glu Gly Glu Glu Val Ala Val Lys Ile Phe
1895 1900 1905

Asn Lys His Thr Ser Leu Arg Leu Leu Arg Gln Glu Leu Val Val
1910 1915 1920

Leu Cys His Leu His His Pro Ser Leu Ile Ser Leu Leu Ala Ala
1925 1930 1935

Gly Ile Arg Pro Arg Met Leu Val Met Glu Leu Ala Ser Lys Gly
1940 1945 1950

Ser Leu Asp Arg Leu Leu Gln Gln Asp Lys Ala Ser Leu Thr Arg
1955 1960 1965

Thr Leu Gln His Arg Ile Ala Leu His Val Ala Asp Gly Leu Arg
1970 1975 1980

Tyr Leu His Ser Ala Met Ile Ile Tyr Arg Asp Leu Lys Pro His
1985 1990 1995

Asn Val Leu Leu Phe Thr Leu Tyr Pro Asn Ala Ala Ile Ile Ala
2000 2005 2010

Lys Ile Ala Asp Tyr Xaa Ile Ala Gln Tyr Cys Cys Arg Met Gly
2015 2020 2025

Ile Lys Thr Ser Glu Gly Thr Pro Gly Phe Arg Ala Pro Glu Val
2030 2035 2040

Ala Arg Gly Asn Val Ile Tyr Asn Gln Gln Ala Asp Val Tyr Ser
2045 2050 2055

Phe Gly Leu Leu Leu Tyr Asp Ile Leu Thr Thr Gly Gly Arg Ile
2060 2065 2070

Val Glu Gly Leu Lys Phe Pro Asn Glu Phe Asp Glu Leu Glu Ile
2075 2080 2085

Gln Gly Lys Leu Pro Asp Pro Val Lys Glu Tyr Gly Cys Ala Pro
2090 2095 2100

Trp Pro Met Val Glu Lys Leu Ile Lys Gln Cys Leu Lys Glu Asn
2105 2110 2115

Pro Gln Glu Arg Pro Thr Ser Ala Gln Val Phe Asp Ile Leu Asn
2120 2125 2130

Ser Ala Glu Leu Val Cys Leu Thr Arg Arg Ile Leu Leu Pro Lys
2135 2140 2145

Asn Val Ile Val Glu Cys Met Val Ala Thr His His Asn Ser Arg
2150 2155 2160

Asn Ala Ser Ile Trp Leu Gly Cys Gly His Thr Asp Arg Gly Gln
2165 2170 2175

Leu Ser Phe Leu Asp Leu Asn Thr Glu Gly Tyr Thr Ser Glu Glu
2180 2185 2190

Val Ala Asp Ser Arg Ile Leu Cys Leu Ala Leu Val His Leu Pro
2195 2200 2205

Val Glu Lys Glu Ser Trp Ile Val Ser Gly Thr Gln Ser Gly Thr
2210 2215 2220

Leu Leu Val Ile Asn Thr Glu Asp Gly Lys Lys Arg His Thr Leu
2225 2230 2235

Glu Lys Met Thr Asp Ser Val Thr Cys Leu Tyr Cys Asn Ser Phe
2240 2245 2250

Ser Lys Gln Ser Lys Gln Lys Asn Phe Leu Leu Val Gly Thr Ala
2255 2260 2265

Asp Gly Lys Leu Ala Ile Phe Glu Asp Lys Thr Val Lys Leu Lys
2270 2275 2280

Gly Ala Ala Pro Leu Lys Ile Leu Asn Ile Gly Asn Val Ser Thr
2285 2290 2295

Pro Leu Met Cys Leu Ser Glu Ser Thr Asn Ser Thr Glu Arg Asn
2300 2305 2310

Val Met Trp Gly Gly Cys Gly Thr Lys Ile Phe Ser Phe Ser Asn
2315 2320 2325

Asp Phe Thr Ile Gln Lys Leu Ile Glu Thr Arg Thr Ser Gln Leu
2330 2335 2340

Phe Ser Tyr Ala Ala Phe Ser Asp Ser Asn Ile Ile Thr Val Val
2345 2350 2355

Val Asp Thr Ala Leu Tyr Ile Ala Lys Gln Asn Ser Pro Val Val
2360 2365 2370

Glu Val Trp Asp Lys Lys Thr Glu Lys Leu Cys Gly Leu Ile Asp
2375 2380 2385

Cys Val His Phe Leu Arg Glu Val Met Val Lys Glu Asn Lys Glu
2390 2395 2400

Ser Lys His Lys Met Ser Tyr Ser Gly Arg Val Lys Thr Leu Cys
2405 2410 2415

Leu Gln Lys Asn Thr Ala Leu Trp Ile Gly Thr Gly Gly Gly His
2420 2425 2430

Ile Leu Leu Leu Asp Leu Ser Thr Arg Arg Leu Ile Arg Val Ile
2435 2440 2445

Tyr Asn Phe Cys Asn Ser Val Arg Val Met Met Thr Ala Gln Leu
2450 2455 2460

Gly Ser Leu Lys Asn Val Met Leu Val Leu Gly Tyr Asn Arg Lys
2465 2470 2475

Asn Thr Glu Gly Thr Gln Lys Gln Lys Glu Ile Gln Ser Cys Leu
2480 2485 2490

Thr Val Trp Asp Ile Asn Leu Pro His Glu Val Gln Asn Leu Glu
2495 2500 2505

Lys His Ile Glu Val Arg Lys Glu Leu Ala Glu Lys Met Arg Arg
2510 2515 2520

Thr Ser Val Glu
2525

<210> 2
<211> 7584
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (6055)..(6055)
<223> n may be a, c or t.

<220>
<221> misc_feature
<222> (5457)..(5457)
<223> n may be c or t.

<220>
<221> misc_feature
<222> (5096)..(5096)
<223> n may be a or g.

<220>
<221> misc_feature
<222> (4321)..(4321)
<223> n may be c or t.

<220>
<221> misc_feature
<222> (3364)..(3364)
<223> n may be a or g.

<220>
<221> misc_feature

<222> (149)..(149)
<223> n may be a or g.

<300>
<308> Genbank / AY792511
<309> 2004-11-15
<313> (1)..(7584)

<400> 2
atggctagtg gcagctgtca ggggtgcgaa gaggacgagg aaactctgaa gaagttgata 60

gtcaggctga acaatgtcca ggaaggaaaa cagatagaaa cgctggtcca aatcctggag 120

gatctgctgg tgttcacgta ctccgagcnc gcctccaagt tatttcaagg caaaaatatc 180

catgtgcctc tgttgatcgt cttggactcc tatatgagag tcgcgagtgt gcagcaggtg 240

ggttggtcac ttctgtgcaa attaatagaa gtctgtccag gtacaatgca aagcttaatg 300

ggaccccagg atgttggaaa tgattgggaa gtccttggtg ttcaccaatt gattcttaaa 360

atgctaacag ttcataatgc cagtgtaaac ttgtcagtga ttggactgaa gaccttagat 420

ctcctcctaa cttcaggtaa aatcaccttg ctgatactgg atgaagaaag tgatattttc 480

atgttaattt ttgatgccat gcactcattt ccagccaatg atgaagtcca gaaacttgga 540

tgcaaagctt tacatgtgct gtttgagaga gtctcagagg agcaactgac tgaatttgtt 600

gagaacaaag attatatgat attgttaagt gcgtcaacaa attttaaaga tgaagaggaa 660

attgtgcttc atgtgctgca ttgtttacat tccctagcga ttccttgcaa taatgtggaa 720

gtcctcatga gtggcaatgt caggtgttat aatattgtgg tggaagctat gaaagcattc 780

cctatgagtg aaagaattca agaagtgagt tgctgtttgc tccataggct tacattaggt 840

aatttttca atatcctggt attaaacgaa gtccatgagt ttgtggtgaa agctgtgcag 900

cagtacccag agaatgcagc attgcagatc tcagcgctca gctgtttggc cctcctcact 960

gagactattt tcttaaatca agatttagag gaaaagaatg agaatcaaga gaatgatgat 1020

gaggggggaag aagataaatt gttttggctg gaagcctgtt acaaagcatt aacgtggcat 1080

agaaagaaca agcacgtgca ggaggccgca tgctgggcac taaataatct ccttatgtac 1140

caaaacagtt tacatgagaa gattggagat gaagatggcc atttcccagc tcatagggaa 1200

gtgatgctct ccatgctgat gcattcttca tcaaaggaag ttttccaggc atctgcgaat 1260

gcattgtcaa ctctcttaga acaaaatgtt aatttcagaa aaatactgtt atcaaaagga 1320

atacacctga atgttttgga gttaatgcag aagcatatac attctcctga agtggctgaa 1380

agtggctgta aaatgctaaa tcatcttttt gaaggaagca acacttccct ggatataatg 1440

gcagcagtgg tccccaaaat actaacagtt atgaaacgtc atgagacatc attaccagtg 1500

cagctggagg cgcttcgagc tattttacat tttatagtgc ctggcatgcc agaagaatcc 1560

agggaggata cagaatttca tcataagcta aatatggtta aaaaacagtg tttcaagaat 1620

gatattcaca aactggtcct agcagctttg aacaggttca ttggaaatcc tgggattcag 1680

aaatgtggat taaaagtaat ttcttctatt gtacattttc ctgatgcatt agagatgtta 1740

tccctggaag gtgctatgga ttcagtgctt cacacactgc agatgtatcc agatgaccaa 1800

gaaattcagt gtctgggttt aagtcttata ggatacttga ttacaaagaa gaatgtgttc 1860

ataggaactg gacatctgct ggcaaaaatt ctggtttcca gcttataccg atttaaggat 1920

gttgctgaaa tacagactaa aggatttcag acaatcttag caatcctcaa attgtcagca 1980

tctttttcta agctgctggt gcatcattca tttgacttag taatattcca tcaaatgtct 2040

tccaatatca tggaacaaaa ggatcaacag tttctaaacc tctgttgcaa gtgttttgca 2100

aaagtagcta tggatgatta cttaaaaaat gtgatgctag agagagcgtg tgatcagaat 2160

aacagcatca tggttgaatg cttgcttcta ttgggagcag atgccaatca agcaaaggag 2220

ggatcttctt taatttgtca ggtatgtgag aaagagagca gtcccaaatt ggtggaactc 2280

ttactgaata gtggatctcg tgaacaagat gtacgaaaag cgttgacgat aagcattggg 2340

aaaggtgaca gccagatcat cagcttgctc ttaaggaggc tggccctgga tgtggccaac 2400

aatagcattt gccttggagg attttgtata ggaaaagttg aaccttcttg gcttggtcct 2460

ttatttccag ataagacttc taatttaagg aaacaaacaa atatagcatc tacactagca 2520

agaatggtga tcagatatca gatgaaaagt gctgtggaag aaggaacagc ctcaggcagc 2580

gatggaaatt tttctgaaga tgtgctgtct aaatttgatg aatggacctt tattcctgac 2640

tcttctatgg acagtgtgtt tgctcaaagt gatgacctgg atagtgaagg aagtgaaggc 2700

tcatttcttg tgaaaaagaa atctaattca attagtgtag gagaatttta ccgagatgcc 2760

gtattacagc gttgctcacc aaatttgcaa agacattcca attccttggg gcccattttt 2820

gatcatgaag atttactgaa gcgaaaaaga aaaatactat cttcagatga ttcactcagg 2880

tcatcaaaac ttcaatccca tatgaggcat tcagacagca tttcttctct ggcttctgag 2940

agagaatata ttacatcact agacctttca gcaaatgaac taagagatat tgatgcccta 3000

agccagaaat gctgtataag tgttcatttg gagcatcttg aaaagctgga gcttcaccag 3060

aatgcactca cgagctttcc acaacagcta tgtgaaactc tgaagagttt gacacatttg 3120

gacttgcaca gtaataaatt tacatcattt ccttcttatt tgttgaaaat gagttgtatt 3180

gctaatcttg atgtctctcg aaatgacatt ggaccctcag tggttttaga tcctacagtg 3240

aaatgtccaa ctctgaaaca gtttaacctg tcatataacc agctgtcttt tgtacctgag 3300

aacctcactg atgtggtaga gaaactggag cagctcattt tagaaggaaa taaaatatca 3360

gggntatgct ccccccttgag actgaaggaa ctgaagattt taaaccttag taagaaccac 3420

atttcatccc tatcagagaa ctttcttgag gcttgtccta aagtggagag tttcagtgcc 3480

agaatgaatt ttcttgctgc tatgcctttc ttgcctcctt ctatgacaat cctaaaatta 3540

tctcagaaca aattttcctg tattccagaa gcaattttaa atcttccaca cttgcggtct 3600

ttagatatga gcagcaatga tattcagtac ctaccaggtc ccgcacactg gaaatctttg 3660

aacttaaggg aactcttatt tagccataat cagatcagca tcttggactt gagtgaaaaa 3720

gcatatttat ggtctagagt agagaaactg catctttctc acaataaact gaaagagatt 3780

cctcctgaga ttggctgtct tgaaaatctg acatctctgg atgtcagtta caacttggaa 3840

ctaagatcct ttcccaatga aatggggaaa ttaagcaaaa tatgggatct tcctttggat 3900

gaactgcatc ttaactttga ttttaaacat ataggatgta aagccaaaga catcataagg 3960

tttcttcaac agcgattaaa aaaggctgtg ccttataacc gaatgaaact tatgattgtg 4020

ggaaatactg ggagtggtaa aaccacctta ttgcagcaat taatgaaaac caagaaatca 4080

gatcttggaa tgcaaagtgc cacagttggc atagatgtga aagactggcc tatccaaata 4140

agagacaaaa gaaagagaga tctcgtccta aatgtgtggg attttgcagg tcgtgaggaa 4200

ttctatagta ctcatcccca ttttatgacg cagcgagcat tgtaccttgc tgtctatgac 4260

ctcagcaagg gacaggctga agttgatgcc atgaagcctt ggctcttcaa tataaaggct 4320

ngcgcttctt cttcccctgt gattctcgtt ggcacacatt tggatgtttc tgatgagaag 4380

caacgcaaag cctgcatgag taaaatcacc aaggaactcc tgaataagcg agggttccct 4440

gccatacgag attaccactt tgtgaatgcc accgaggaat ctgatgcttt ggcaaaactt 4500

cggaaaacca tcataaacga gagccttaat ttcaagatcc gagatcagct tgttgttgga 4560

cagctgattc cagactgcta tgtagaactt gaaaaaatca ttttatcgga gcgtaaaaat 4620

gtgccaattg aatttcccgt aattgaccgg aaacgattat tacaactagt gagagaaaat 4680

cagctgcagt tagatgaaaa tgagcttcct cacgcagttc actttctaaa tgaatcagga 4740

gtccttcttc attttcaaga cccagcactg cagttaagtg acttgtactt tgtggaaccc 4800

aagtggcttt gtaaaatcat ggcacagatt ttgacagtga aagtggaagg ttgtccaaaa 4860

caccctaagg gcattatttc gcgtagagat gtggaaaaat ttctttcaaa aaaaaggaaa 4920

tttccaaaga actacatgtc acagtatttt aagctcctag aaaaattcca gattgctttg 4980

ccaataggag aagaatattt gctggttcca agcagtttgt ctgaccacag gcctgtgata 5040

gagcttcccc attgtgagaa ctctgaaatt atcatccgac tatatgaaat gccttntttt 5100

ccaatgggat tttggtcaag attaatcaat cgattacttg agatttcacc ttacatgctt 5160

tcagggagag aacgagcact tcgcccaaac agaatgtatt ggcgacaagg catttactta 5220

aattggtctc ctgaagctta ttgtctggta ggatctgaag tcttagacaa tcatccagag 5280

agtttcttaa aaattacagt tccttcttgt agaaaaggct gtattctttt gggccaagtt 5340

gtggaccaca ttgattctct catggaagaa tggtttcctg ggttgctgga gattgatatt 5400

tgtggtgaag gagaaactct gttgaagaaa tgggcattat atagttttaa tgatggngaa 5460

gaacatcaaa aaatcttact tgatgacttg atgaagaaag cagaggaagg agatctctta 5520

gtaaatccag atcaaccaag gctcaccatt ccaatatctc agattgcccc tgacttgatt 5580

ttggctgacc tgcctagaaa tattatgttg aataatgatg agttggaatt tgaacaagct 5640

ccagagtttc tcctaggtga tggcagtttt ggatcagttt accgagcagc ctatgaagga 5700

gaagaagtgg ctgtgaagat ttttaataaa catacatcac tcaggctgtt aagacaagag 5760

cttgtggtgc tttgccacct ccaccacccc agtttgatat ctttgctggc agctgggatt 5820

cgtccccgga tgttggtgat ggagttagcc tccaagggtt ccttggatcg cctgcttcag 5880

caggacaaag ccagcctcac tagaacccta cagcacagga ttgcactcca cgtagctgat 5940

ggtttgagat acctccactc agccatgatt atataccgag acctgaaacc ccacaatgtg 6000

ctgcttttca cactgtatcc caatgctgcc atcattgcaa agattgctga ctacngcatt 6060

gctcagtact gctgtagaat ggggataaaa acatcagagg gcacaccagg gtttcgtgca 6120

cctgaagttg ccagaggaaa tgtcatttat aaccaacagg ctgatgttta ttcatttggt 6180

ttactactct atgacatttt gacaactgga ggtagaatag tagagggttt gaagtttcca 6240

aatgagtttg atgaattaga aatacaagga aaattacctg atccagttaa agaatatggt 6300

tgtgccccat ggcctatggt tgagaaatta attaaacagt gtttgaaaga aaatcctcaa 6360

gaaaggccta cttctgccca ggtctttgac attttgaatt cagctgaatt agtctgtctg 6420

acgagacgca ttttattacc taaaaacgta attgttgaat gcatggttgc tacacatcac 6480

aacagcagga atgcaagcat ttggctgggc tgtgggcaca ccgacagagg acagctctca 6540

tttcttgact taaatactga aggatacact tctgaggaag ttgctgatag tagaatattg 6600

tgcttagcct tggtgcatct tcctgttgaa aaggaaagct ggattgtgtc tgggacacag 6660

tctggtactc tcctggtcat caataccgaa gatgggaaaa agagacatac cctagaaaag 6720

atgactgatt ctgtcacttg tttgtattgc aattcctttt ccaagcaaag caaacaaaaa 6780

aattttcttt tggttggaac cgctgatggc aagttagcaa tttttgaaga taagactgtt 6840

aagcttaaag gagctgctcc tttgaagata ctaaatatag gaaatgtcag tactccattg 6900

atgtgtttga gtgaatccac aaattcaacg gaaagaaatg taatgtgggg aggatgtggc 6960

acaaagattt tctccttttc taatgatttc accattcaga aactcattga gacaagaaca 7020

agccaactgt tttcttatgc agctttcagt gattccaaca tcataacagt ggtggtagac 7080

actgctctct atattgctaa gcaaaatagc cctgttgtgg aagtgtggga taagaaaact 7140

gaaaaactct gtggactaat agactgcgtg cacttttaa gggaggtaat ggtaaaagaa 7200

aacaaggaat caaaacacaa aatgtcttat tctgggagag tgaaaaccct ctgccttcag 7260

aagaacactg ctctttggat aggaactgga ggaggccata ttttactcct ggatctttca 7320

actcgtcgac ttatacgtgt aatttacaac ttttgtaatt cggtcagagt catgatgaca 7380

gcacagctag gaagccttaa aaatgtcatg ctggtattgg gctacaaccg gaaaaatact 7440

gaaggtacac aaaagcagaa agagatacaa tcttgcttga ccgtttggga catcaatctt 7500

ccacatgaag tgcaaaattt agaaaaacac attgaagtga gaaaagaatt agctgaaaaa 7560

atgagacgaa catctgttga gtaa 7584

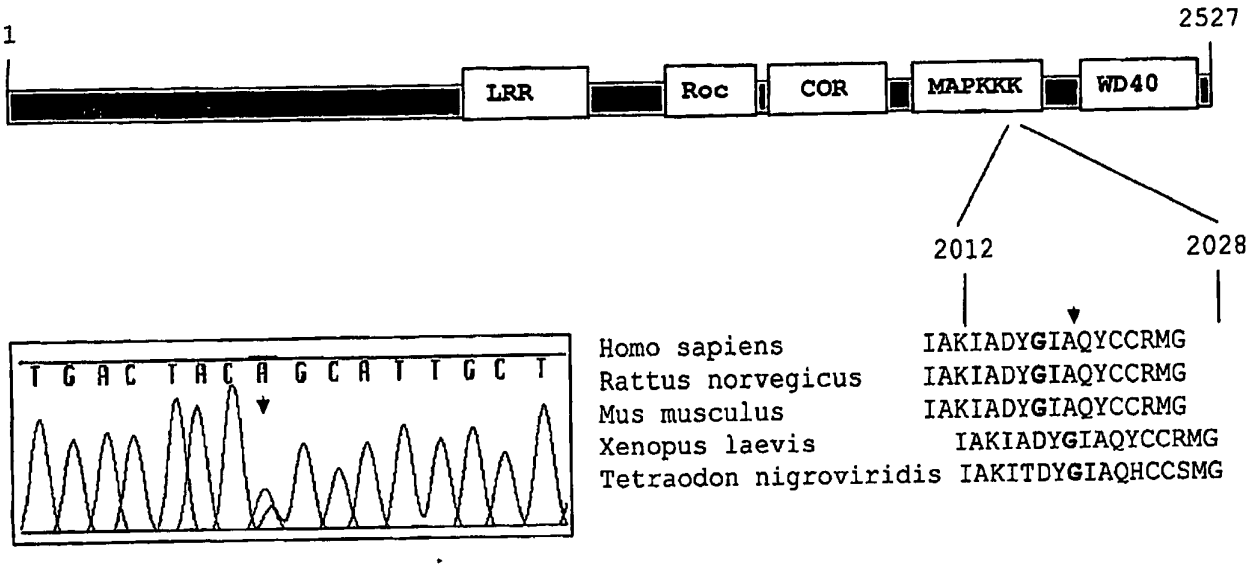
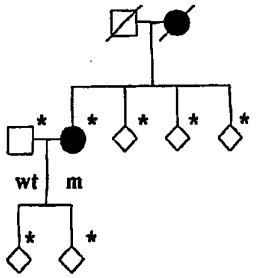


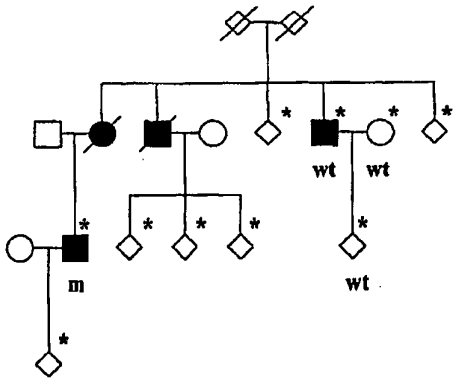

**Figure 1. Schematic drawing of *LRRK2* with predicted protein domains**

(LRR – leucine rich repeat, Roc – Ras in complex proteins, COR – domain C-terminal of Roc, MAPKKK – mitogen-activated protein kinase kinase kinase, WD40 – WD40 repeats). The human *LRRK2* protein sequence in the region of the G2019S mutation is aligned with orthologs from rat (XP_235581), mouse (AAH34074), frog (AAH76853), and puffer fish (CAG05593). The chromatogram shows the 6055G>A transition (G2019S)

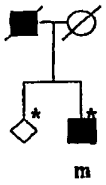
Family P-063



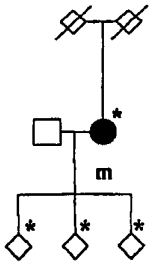

Family P-089



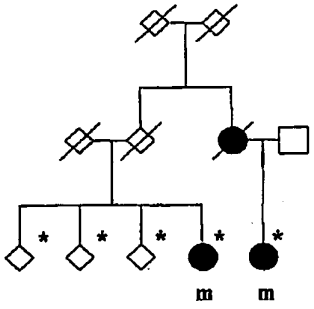

Family P-104



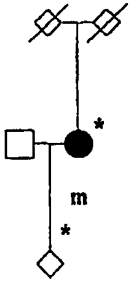

Family P-241



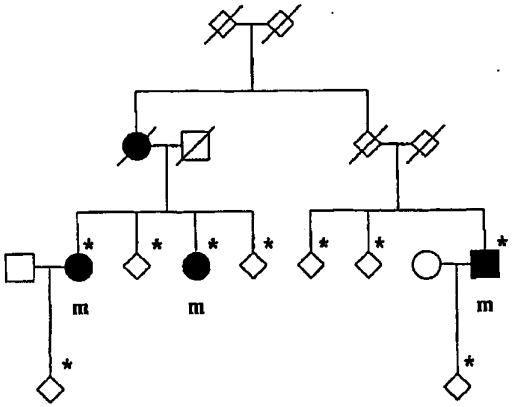

Family P-369



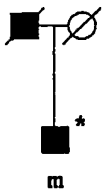

Family P-394




Family F05




Family 1210



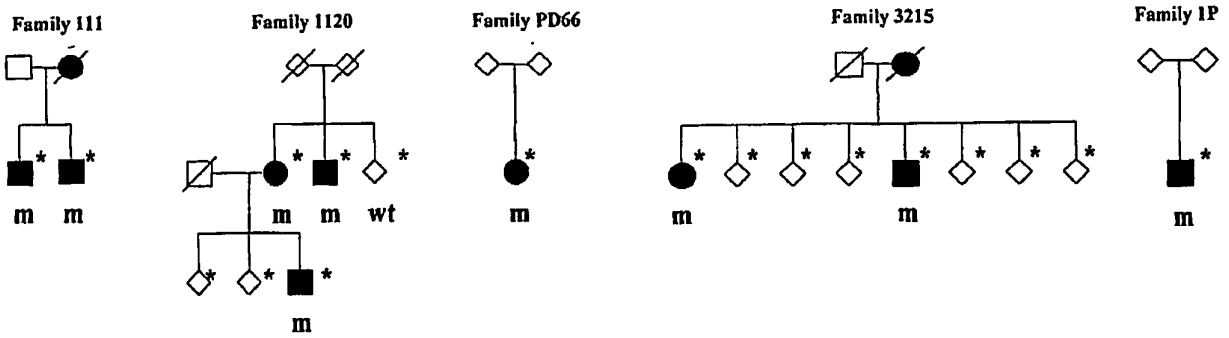




**Figure 2. Pedigrees of families with *LRRK2* G2019S**

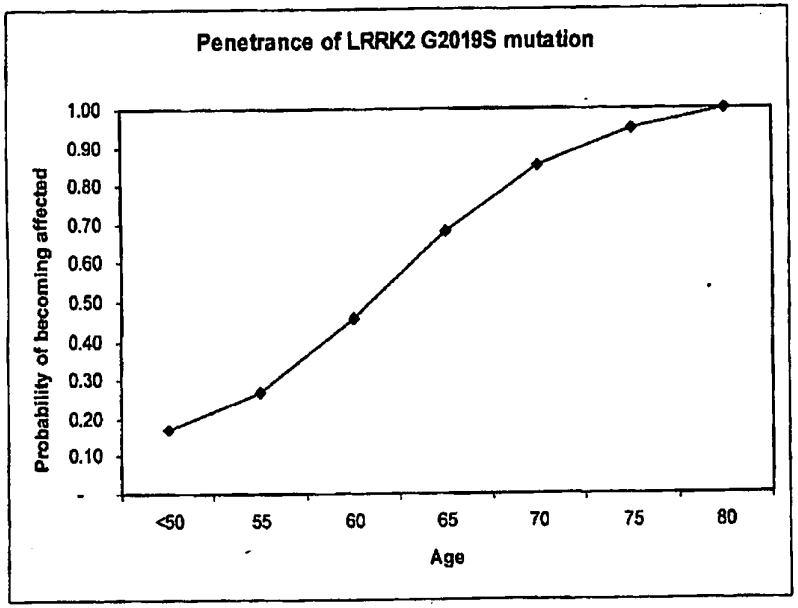
□ and○ denotes sexes, and ◊ denotes that the sex is not given. A diagonal line across the symbol denotes that the person is dead, and thus that he/she has not been tested. Blackened symbols denote affected family members with parkinsonism. An asterisk denotes genotyped individual, with "m" for mutation carriers and "wt" for wild-type *LRRK2*. To protect confidentiality, the genotypes and genders of some unaffected individuals are not shown.

**Figure 3. Chromosome 12q12 STR markers on the disease haplotype (PARK8).**

| | Family proband | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Marker** | **P-063** | **P-089** | **P-104** | **P-241** | **P-369** | **P-394** | **F05** | **1210** | **1120** | **111** | **3215** | **PD66** | **1P** |
| **D12S87** | 160 | 160 | 164 | 164 | - | 156 | 166 | 156/158 | 164 | 160 | 158 | 156/166 | 156/158 |
| **D12S1648** | 120 | 120 | 122 | 122 | 122 | 110 | 110 | 122/124 | 110 | 110 | 110 | 120/134 | 128/130 |
| **D12S2080** | 188 | 188 | 188 | 188 | 188 | 188 | 188 | 184/192 | 188 | 180 | 184 | 188/192 | 184/188 |
| **D12S2194** | 265 | 265 | 265 | 265 | 265 | 265 | 261 | 253/261 | 257 | 257 | 253 | 245/249 | 249/261 |
| **-31Kb** | 290 | 290 | 290 | 290 | 290 | 290 | 290 | 290/290 | 290 | 290 | 290 | 290/293 | 284/290 |
| **LRRK2_69Kb** | 223 | 223 | 223 | 223 | 223 | 223 | 223 | 219/223 | 223 | 223 | 223 | 215/215 | 211/219 |
| **LRRK2_84Kb** | 253 | 253 | 253 | 253 | 253 | 253 | 253 | 253/253 | 253 | 253 | 253 | 253/253 | 253/253 |
| **LRRK2_129Kb** | 151 | 151 | 151 | 151 | 151 | 151 | 151 | 151/151 | 151 | 151 | 151 | 151/151 | 151/151 |
| **212Kb** | 132 | 132 | 132 | 132 | 132 | 132 | 132 | 132/132 | 132 | 132 | 132/138 | 132/138 | 132/134 |
| **243Kb** | 315 | 315 | 315 | 315 | 315 | 315 | 315 | 315/315 | 315 | 315 | 315/309 | 315/312 | 315/300 |
| **378Kb** | 189 | 189 | 189 | 189 | 189 | 189 | 189 | 189/193 | 193 | 193 | 191 | 183/189 | 183/187 |
| **D12S1048** | 214 | 214 | 214 | 214 | 214 | 211/214 | 214 | 214/223 | 214 | 214 | 223 | 211/214 | 211/226 |
| **D12S1301** | 112 | 116 | 120 | 120 | 116 | 116 | 116 | 108/116 | 100 | 120 | 116 | 100/116 | 100/100 |
| **D12S1701** | 95 | 97 | 91 | 91 | 95 | 95/97 | 97 | 95/101 | 92 | 91/95 | 95 | 97/101 | 91/97 |
| **Country of origin** | **Norway** | | | | | | | **United States** | | | **Ireland** | | **Poland** |

Genotypes for probands from13 families with *LRRK2* G2019S are shown, those shared are highlighted in grey.

**Figure 4. Probability of becoming affected by parkinsonism, in *LRRK2*G2019S carriers, as a function of age.**



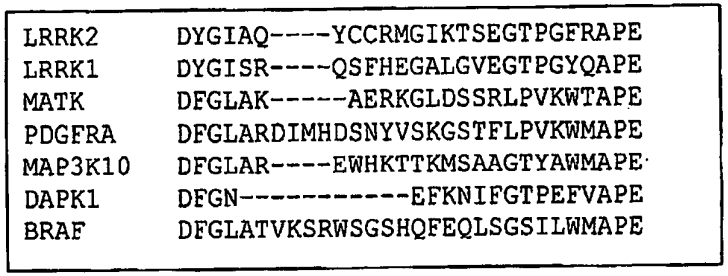



**Figure 5. Aligned amino acid sequences of the activation loop of different human kinases.**

In most kinases, the activation loop starts and ends with the conserved residues DFG and APE, respectively . In *LRRK2* and *LRRK1* phenylalanine is changed to tyrosine, an amino acid with a similar structure. (LRRK2 – leucine-rich repeat kinase 2, LRRK1 – leucine-rich repeat kinase 1, MATK – megakaryocyte-associated tyrosine kinase, PDGFRA – platelet-derived growth factor receptor alpha, MAP3K10 – mitogen-activated protein kinase kinase kinase 10, DAPK1 – death-associated protein kinase 1, BRAF – v-raf murine sarcoma viral oncogene homolog B1)